



TITLE:

# Sound imaging of nocturnal animal calls in their natural habitat.

AUTHOR(S):

Mizumoto, Takeshi; Aihara, Ikkyu; Otsuka, Takuma;  
Takeda, Ryu; Aihara, Kazuyuki; Okuno, Hiroshi G

---

CITATION:

Mizumoto, Takeshi ...[et al]. Sound imaging of nocturnal animal calls in their natural habitat.. Journal of comparative physiology. A, Neuroethology, sensory, neural, and behavioral physiology 2011, 197(9): 915-921

ISSUE DATE:

2011-09

URL:

<http://hdl.handle.net/2433/147258>

RIGHT:

The final publication is available at [www.springerlink.com](http://www.springerlink.com); This is not the published version. Please cite only the published version.; この論文は出版社版ではありません。引用の際には出版社版をご確認ご利用ください。

Journal of Comparative Physiology A manuscript No.  
(will be inserted by the editor)

## Sound Imaging of Nocturnal Animal Calls in Their Natural Habitat

Takeshi Mizumoto · Ikkyu Aihara ·  
Takuma Otsuka · Ryu Takeda · Kazuyuki Aihara ·  
Hiroshi G. Okuno

Received: date / Accepted: date

**Abstract** We present a novel method for imaging acoustic communication between nocturnal animals. Investigating the spatio-temporal calling behavior of nocturnal animals, e.g., frogs and crickets, has been difficult because of the need to distinguish many animals' calls in noisy environments without being able to see them. Our method visualizes the spatial and temporal dynamics using dozens of sound-to-light conversion devices (called "Firefly") and an off-the-shelf video camera. Firefly consisting of a microphone and a light emitting diode emits light when it captures nearby sound. Deploying dozens of Fireflies in a target area, we record calls of multiple individuals through the video camera. We conduct two experiments, one indoors and the other in the field, using Japanese tree frogs (*Hyla japonica*). The indoor experiment demonstrates that our method correctly visualizes Japanese tree frogs' calling behavior. It has confirmed the known behavior; two frogs call synchronously or in anti-phase synchronization. The field experiment (in a rice paddy where Japanese tree frogs live) also visualizes the same calling behavior to confirm anti-phase synchronization in the field. Experimental results confirm that our method can visualize the calling behavior of nocturnal animals in their natural habitat.

**Keywords** Sound Imaging; Visualization; Acoustic Communication; Nocturnal Animal; Measurement Method

---

T. Mizumoto  
Engineering Building #10, Graduate School of Informatics, Kyoto University, Sakyo, Kyoto 606-8501, Japan.  
Tel: +81-75-753-4962  
Fax: +81-75-753-5977  
E-mail: mizumoto@kuis.kyoto-u.ac.jp

I. Aihara  
Department of Physics, Graduate School of Sciences, Kyoto University Kyoto 606-8502, Japan.

T. Otsuka, R. Takeda and H. G. Okuno  
Engineering Building #10, Graduate School of Informatics, Kyoto University, Sakyo, Kyoto 606-8501, Japan.

K. Aihara  
Institute of Industrial Science, University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo 153-8505, Japan.

## Introduction

Acoustic communication plays important roles in the lives of nocturnal animals (Gerhardt and Huber 2002; Wells 2007) because auditory information is more reliable than visual information at night. The communication is observed in not a few kinds of animals, for example, frogs (Feng et al. 2006), crickets (Hedwig and Poulet 2004), and bats (Riquimaroux et al. 1991). In particular, the spontaneous calling behavior of many individual animals in the wild, e.g., a chorus of frogs, has been an intriguing research topic for a long time (Narins and Capranica 1978; Simmons 2004; Suggs and Simmons 2005). The behavior is thought to be a means of maintaining territory, making it easier to be distinguished by mates (see (Gerhardt and Huber 2002; Wells 2007; Aihara 2009) for details.) However, experimental investigation of nocturnal behavior has been difficult because of (1) the co-existence of various kinds of sounds, which makes it difficult to distinguish each individual's calls and (2) darkness, which prevents researchers from directly watching the behavior.

This paper presents a visualization method of the spatio-temporal calling behavior of multiple nocturnal animals in their natural habitat, named *sound imaging*. The method uses newly-developed devices for converting input sound into light, which are distributed in a target area to detect the spatio-temporal dynamics of animal calling behavior. Generally, spatio-temporal calling behavior is observed through the following three steps: (1) visualizing the behavior, (2) estimating the time and location of each call, and (3) separating the calls from a mixture of sounds. In this paper, we propose solutions to the first two problems. The method estimates only the time and location of an animal's call. Nevertheless, it is expected to reveal important aspects of acoustic communication. For example, based on the observation of the timing of calls in an indoor environment, we are developing a communication model of Japanese tree frogs (*Hyla japonica*) using a mathematical model of coupled oscillators (Aihara 2009; Aihara et al. 2011).

Many methods for sound source localization and separation from a mixture of sounds have been developed only for indoor use. For example, blind source separation using independent component analysis (Hyvarinen et al. 2001; Sawada et al. 2003) separates the sound mixture without any prior information but the observed signal. A widely used sound source localization method, MUSIC (Multiple Signal Classification) (Asano et al. 1999), and a geometrically constrained higher-order decorrelation-based source separation method (GHDSS) (Nakajima et al. 2010) are incorporated in the open-source software, HARK (Nakadai et al. 2009).

A system for investigating the spatio-temporal behavior of a chorus, i.e., successive calls of many animals of same species, is still an open problem although many researchers have tried to localize sounds in the field. MacCurdy and Fristrup (2009) investigated the locations of American crows by attaching loggers to them. Such a logger cannot be used for Japanese tree frogs because it is too large to attach. Schwartz (2001) estimated frog location by comparing sound power captured by microphones with a threshold power. Because the method selects the microphone closest to the frog, the spatial resolution is limited by the number of microphones. The localization method using arrival-time differences between microphones is difficult to apply to a dense chorus. This is because the method assumes that the beginning and ending time of sources in which at most one single source exists are given in advance. This kind of method is widely-used for such animals as marine animals (Spiesberger and Fristrup 1990), lek-breeding reed frogs (Grafe 1997), red-winged blackbirds (Spiesberger 1999), bowhead whales (Clark and Ellison 2000), and bullfrogs (Simmons et al. 2008). Although Jones and Ratnam have developed both localization and separation methods and used them to investigate the acoustic communication of two kinds of animals: Gulf Coast toads

and Northern cricket frogs (Jones and Ratnam 2009), they also adopted the same assumption of temporal sparseness, that is, at most only one source exists at the same time.

We present an *inexpensive and easy-to-use* method for bioacoustic investigations. Our solution is a sound imaging system consisting of dozens of “Firefly” sound imaging devices and an off-the-shelf video camera. The Firefly is composed of a microphone, an amplifier, and a light-emitting diode (LED). Nearby sounds captured by the Firefly are converted into light. The sound imaging system follows the procedure: Fireflies are placed on the ground around the sound sources, and their light emissions are recorded with a video camera. The recorded emissions are analyzed to obtain a visualization of the animal spatio-temporal calling behaviors. This visualization enables us to estimate when and from where the animals called.

We evaluated our sound imaging system both indoors and in the field for a particular nocturnal calling animal, namely Japanese tree frog. The indoor experiment was a basic evaluation to determine whether the system was capable of visualizing the spatio-temporal behavior of sound sources. The indoor experimental environment was ideal in the sense that (1) reference data was available, i.e., we knew the frogs’ locations and can record their calls with microphones, which is difficult in the field. (2) Only the sounds of interest occurred and (3) the video camera could be placed orthogonally to the target area without obstacles. Thus, we could record the Firefly emissions with smaller lens distortion and without occlusion. The field experiment was more challenging: (1) No prior information was available about when and from where the frogs call, (2) the field was noisy because of the presence of other species and other Japanese tree frogs outside the target area, and (3) the video camera was panned and tilted in order to cover the field of interest, which may cause severe lens distortion.

## Materials and methods

### Data acquisition equipment

We use an off-the-shelf video camera (Sony Corp., HDR-XR550V, 29.97 fps) and our newly designed sound imaging device, Firefly. It consists of a microphone, an LED, and an amplifier with a gain that can be adjusted using a variable resistor (see Electronic Supplementary Material for a circuit diagram of Firefly, a picture of Firefly, and the list of electronic parts. Note that all parts are commercially available.) The circuit consists of four modules: a power source, a microphone, an amplifier, and a sound-to-light converter. The power source module generates two voltages, 2.4V and 5V, for the microphone and amplifier, respectively. The microphone module, which consists of an electret condenser microphone and its peripheral circuit, captures nearby sound. Then, the amplifier module increases the amplitude of the sound’s waveform. Finally, the sound-to-light conversion module drives the LED using the rectified waveform.

### Data analysis for sound imaging

Visualizing the light pattern corresponding to the spatio-temporal behavior of frogs requires two functions: (1) detecting a weak light emitted by each small device (2) eliminating the individual differences among devices. The key ideas to fulfill these requirements are (1) increasing the contrast between the emitted lights and the background using frame averaging

and (2) subtracting its mean from each light intensity time series. The procedure is described as follows.

(I) First, we divide the video file into individual frames using a movie processing tool, TMPGEnc by Pegasys Inc. Let the pixel values of the  $t$ th colored frame be  $I(x, y, t)$ , where  $x$  and  $y$  denote the orthogonal coordinates of the frame. Let the red, green, and blue components of the frame be  $I_R(x, y, t)$ ,  $I_G(x, y, t)$ , and  $I_B(x, y, t)$ , respectively. Then, we convert each colored frame to the gray-scaled frame  $I_g(x, y, t)$  using the following equation defined in the NTSC (National Television System Committee) Standard for calculating brightness from a color pixel (ITU-R 2007):

$$I_g(x, y, t) = 0.2989 \times I_R(x, y, t) + 0.5870 \times I_G(x, y, t) + 0.1140 \times I_B(x, y, t).$$

(II) We then detect the positions of the LEDs in the frame and generate masks to cover them. Since we assume that the devices and the camera do not move during the experiment, we first calculate the mean frame by averaging  $I_g$  as follows:

$$I_{mean}(x, y) = \frac{1}{N} \sum_{t=1}^N I_g(x, y, t),$$

where  $N$  denotes the number of frames. We convert  $I_{mean}$  to a binary frame by comparing it with a threshold determined using discriminant analysis (Otsu 1979). We then segment the binary image based on 4-connectivity, which is a segmentation criterion in which each pixel  $(x, y)$  is connected to four pixels  $(x \pm 1, y)$  and  $(x, y \pm 1)$ . Each  $m$ th segment  $I_m(x, y)$  corresponds to the mask for the LED of each device.

$$I_m(x, y) = \begin{cases} 1, & \text{if } I_{mean}(x, y) \text{ belongs to } m\text{-th LED} \\ 0, & \text{otherwise} \end{cases}$$

where  $m (= 1, \dots, M)$  denotes the mask index. The number of masks,  $M$ , is equal to the number of detected devices. We used the `bwlabel(BW, n)` function in the MATLAB Image Processing Toolbox to achieve this segmentation. The `bwlabel` with  $n = 4$  returns labels of 4-connected components in 2-D binary images. We then manually eliminate invalid segments on the basis of their area and location.

(III) Next, we extract the time series  $t_s(m, t)$  of each device's light intensity by using the masks:

$$t_s(m, t) = \sum_{x=0}^W \sum_{y=0}^H \left( I_m(x, y) I_g(x, y, t) \right),$$

where  $W$  and  $H$  denote the frame width and height, respectively.

(IV) We then eliminate the errors due to the light intensity variations among the Firefly devices (caused by the tolerance of the integrated circuits, resistors, and transistors) by subtracting the average of  $t_s(m, t)$  for each mask:

$$\hat{t}_s(m, t) = t_s(m, t) - \frac{1}{T} \sum_{t=1}^T t_s(m, t).$$

We assume that these errors are additive. The resulting time series  $\hat{t}_s(m, t)$  describes the spatio-temporal calling behavior.

(V) Then, using a nonlinear process called gamma correction, we decipher the extracted time series. Extracting it directly is difficult because the contrast of visualization may be low. Gamma correction is commonly used to correct contrast in image processing. The  $\gamma$ th

power,  $\hat{t}_s(t, m)^\gamma$ , is calculated for each pixel, where  $\gamma$  is a real number (e.g., see (Farid 2001) for details). In the analysis, we used  $\gamma = 2.5$ .

(VI) Finally, we estimate the time and location of each call by assuming that a peak in the time series corresponds to a call. In preprocessing, we set any component less than the threshold  $\alpha$  to zero to eliminate noise; in the analysis, we empirically set  $\alpha = 0.1$ .

The peaks are detected by sliding a square window of size  $L$  to extract a part of the time series  $\hat{t}_s$ . If the center of the part is the maximum, the position corresponds to a call. This algorithm is formulated as follows: a call is made at the location  $m$  and the time  $t$  when  $m$  and  $t$  satisfy

$$\arg \max_{\hat{m}, \hat{t}} \hat{t}_s(\hat{m}, \hat{t})$$

where  $\hat{m} \in [m - L, \dots, m + L]$  and  $\hat{t} \in [t - L, \dots, t + L]$ .

## Experiments

### *Indoor experiment*

The indoor experiment was conducted on 8 June 2009. Two Japanese tree frogs were caught in the field of the experimental farm of Kyoto University, Japan, and placed in separate cages. The snout-vent lengths were 38.83mm and 32.85mm, and their respective weights were 2.4g and 2.5g.

In the experiment room, we positioned 20 Fireflies in two rows at intervals of 21.5cm and installed an off-the-shelf video camera above the experimental area at a height of 5.13m. Then, we placed one frog between the 3rd and 4th Fireflies and the other one between the 7th and 8th Fireflies (see Fig. 1a). The humidity was 54% and the temperature was 25.0°C. We recorded the Firefly emissions throughout one night.

### *Field experiment*

The field experiment was conducted alongside a rice paddy field in Oki Island, Shimane Prefecture, Japan, on 3 July 2009. The humidity was 76%, and the temperature was 24°C.

We placed 20 Fireflies along the edge of the paddy, and the off-the-shelf video camera on the side of the paddy (see Fig. 1b). Then, we recorded the Firefly emissions for one hour at night. We neither caught any frogs nor identified any frog locations because the goal was to record the frogs' behavior in their natural habitat.

## Results

### *Indoor experiment*

The two peaks in the sound imaging results (Fig. 2a shows an excerpt of the result) around the 3rd and 7th Fireflies correspond to the frogs' locations. This means that our method correctly visualize the sound source locations. The magnification in Fig. 2b for 15.0 to 17.0s clearly shows the sequences for the two peaks. The alternation of the peaks means that the two frogs called alternately. This anti-phase synchronization behavior is the same as that reported by Aihara (2009); Aihara et al. (2011).

Figure 2c shows the waveforms recorded for the calls of the two frogs. They were recorded using two microphones placed near the frogs to enable us to evaluate the temporal accuracy of our system. Because their calls were loud, the recorded waveforms were a mixture of the two frogs' calls. We separated them using an independent component analysis (Hyvarinen et al. 2001; Sawada et al. 2003). Note that the ICA is applicable to this case, because (1) the number of sound sources is the same as that of microphones and (2) the amplitude histogram of Japanese tree frog calls is non-Gaussian (Aihara et al. 2011). The waveform in the upper panel in Fig. 2c corresponds to the sequence of peaks for the 3rd Firefly, and the lower one corresponds to that for the 7th Firefly. Calling detections illustrated as white pixels (the lower panel of Fig. 2b) coincide with the peaks of the waveform (Fig. 2c). This means that our method correctly visualize the temporal aspect of the calling behaviors.

These results demonstrate that our method is capable of visualizing the spatio-temporal calling behavior, i.e., both the call times and locations, of Japanese tree frogs.

### Field experiment

In the field experiment, we recorded 20 frog-calling sequences. The average number of calls per sequence was 238. Figure 3 shows a histogram of the durations.

As visualized in Fig. 2d, two frogs near the 7th and 11th Fireflies started calling at 2.5s and ended at 15.0s. The sequence of the peaks for their call from 5.0s to 7.0s (Fig. 2e) shows that the two frogs called alternately, i.e., in anti-phase synchronization. This result confirms that two Japanese tree frogs call in anti-phase synchronization even in their natural habitat. We thus succeeded in observing the calling behavior of Japanese tree frogs in their natural habitat with high temporal (1/30s) and spatial resolutions (21.5cm indoors and 31.6cm in the field).

### Localization

The lower panels in Figs. 2b and 2e show the times of the calls and the locations of the two frogs estimated from the corresponding upper panels. The average locations of the frogs were 131.2cm and 61.3cm for the indoor experiment and 221.1cm and 315.8cm for the field experiment. The standard deviations were 7.1cm and 7.6cm for the indoor experiment and both zero for the field experiment. These values were calculated by multiplying the device index by the Firefly intervals (21.5cm for the indoor experiment and 31.6cm for the field experiment).

We calculated the estimation error for the indoor experiment because the locations of the frogs were known. They were near the 3rd and 7th Fireflies, which correspond to 64.5cm and 150.5cm, respectively. Therefore, the mean estimation errors were 3.2cm and 19.3cm, respectively. These errors are less than the device interval.

Although the locations of the frogs were unknown in the field experiment, we evaluate the distance between them. The distance was calculated by multiplying the device index by the intervals to be 102.6cm. This is consistent both with the finding of a previous study that Japanese tree frogs are sparsely distributed in a field (Matsui 1996; Maeda and Matsui 1999), and with our empirical observation that they call at a distance of 1m to 3m.

## Discussion

We proposed a novel sound imaging system and showed that it is capable of localizing calling frogs. Our results confirm the previously-observed synchronized calling of Japanese tree frogs in the field (Aihara 2009; Aihara et al. 2011). The behavior is a biologically nontrivial phenomenon because methods based on signal processing techniques often fail due to the presence of many kinds of sound sources in the field in addition to the target sounds.

The simplicity and low cost of our system are the key advantages for use in bioacoustic studies. The acoustic communication can be observed by simply distributing Fireflies and recording their emissions. The system can easily be scaled up to cover a larger area, by simply adding more Fireflies and video cameras. In our experience, when an ordinary microphone, e.g., the electret condenser microphone ECM-C10 produced by Sony Corp., is used, it detects distant environmental noise such as calls of other frogs in a paddy field that are out of our interest due to its sensitivity to distant sound sources. In contrast, our system is less affected by such noise because the gain of the microphone used for Firefly decays quickly with distance. We roughly calibrate the sensitivity of each Firefly so that it emits light when a frog calls within a radius of about 30cm.

Our sound imaging system has several qualitative advantages over existing ones, such as attached loggers and microphone arrays. As mentioned above, attaching loggers is not acceptable for small animals such as Japanese tree frogs. In microphone-array methods, the calculated arrival-time differences or the correlation are typically used to estimate the location of the sound. These methods implicitly assume sparseness in the time-frequency domain, i.e., only one call at a time at each frequency. Usually, the assumption does not hold for the dense chorus of animals because animals of the same species tend to have calls of similar frequency characteristics. Although Jones and Ratnam (2009) were able to find in their recordings time intervals during which only one frog called, such recordings need a lot of time and efforts. In contrast, our system identifies the calling behavior automatically, and thus it is easier to use than other methods.

The performance of the system when there are multiple overlapping callers is important especially for observations in the field because the overlapping can easily occur in a dense chorus. There are two kinds of overlapping, temporal and spatial. The temporal overlapping is not crucial if spatial sparseness holds because the calls appear at different locations in the light pattern. Figure 4 shows an example of the system successfully localizing the overlapping calls of two frogs.

The spatial overlapping, e.g., two calling animals near each other, is still a crucial problem. When two animals are present closer together than the Firefly intervals, the system cannot distinguish their calls because the lighting pattern is the same for both animals' calls. This problem can be mitigated by reducing the Firefly intervals so that the spatial resolution improves with the Firefly density. Because of the scalable spatial resolution of our system, it can be applied to other kinds of animal choruses with different spatio-temporal densities. Even if the target animal's chorus is spatially denser than that of Japanese tree frogs, the system can visualize the chorus with more Fireflies.

Future work includes evaluating the spatial and temporal accuracy in detail. By annotating call times of the recorded data from the indoor experiments, we can evaluate the temporal accuracy of the current system. The spatial accuracy can be evaluated using loudspeakers placed at various distances. This evaluation will reveal the relationship between spatial density of sound sources and localization error.



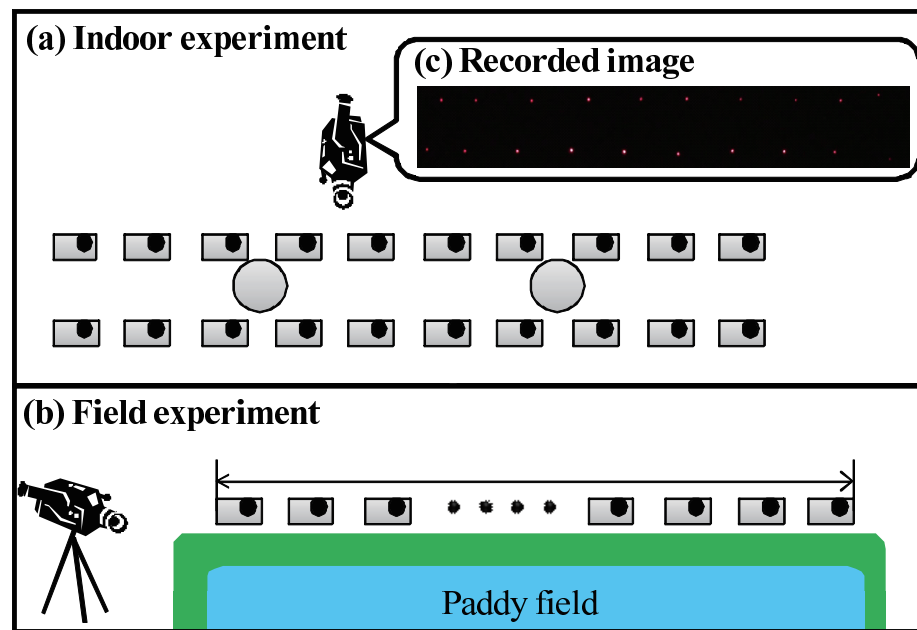
We plan to enhance the system in two ways. One is to downsize the device by using chip parts such as MEMS (micro electro mechanical systems) microphones (Liu 2005). The smaller the device, the easier it is to carry and deploy in the field. The other is to implement modular sound sensors and LEDs so that the device can easily adapt to the frequency range of the calls of the target species (Riquimaroux et al. 1991; Gerhardt and Huber 2002; Hedwig and Poulet 2004; Feng et al. 2006; Wells 2007) as well as to its visual spectrum. For example, the microphones can be replaced with ultrasonic sensor for the frogs that communicate using ultrasonic sounds (Feng et al. 2006). The LEDs can be replaced with infrared LEDs for animals that do not see infrared. Even with such changes, the framework of our sound imaging system will remain effective. The automatic estimation of parameters, such as  $\alpha$  and  $\gamma$ , is also important when we apply our system to various species.

**Acknowledgements** This work was supported by the JSPS Grant-in-Aid for Exploratory Research (No. 21650043), the JSPS Grant-in-Aid for (S) (No. 19100003), the JSPS Grant-in-Aid for JSPS Fellows (No. 08J00608), the JSPS FIRST Program, and Honda Research Institute Japan, Co. Ltd. We would like to thank T. Kobayashi and H. Kitahata for their suggestions for analyzing the recorded emission data, H. Riquimaroux, P. M. Narins, K. Okanoya and A. Yamaguchi for their helpful advice, and A. Lim and L. K. Cahier for suggestions to improve the English of this manuscript. We performed all experiments in accordance with the guidelines of the Animal Research Committee of Kyoto University.

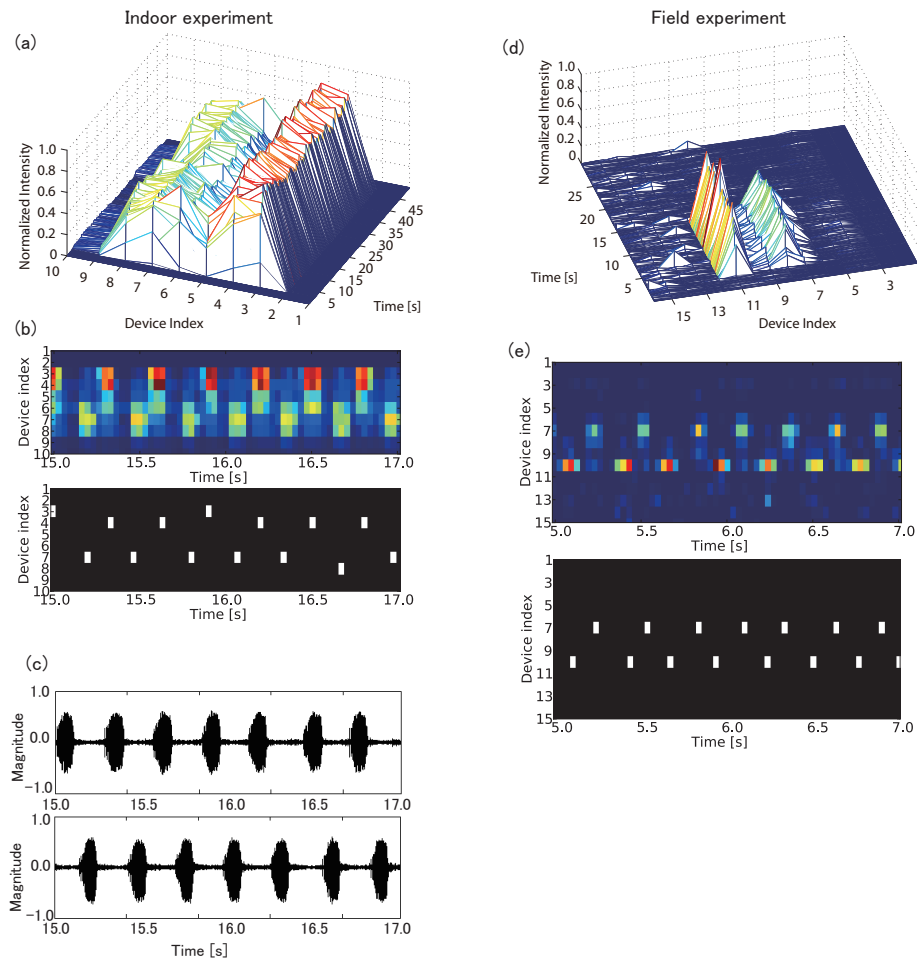
## References

- Aihara I (2009) Modeling synchronized calling behavior of Japanese tree frogs. *Phys Rev E* 8:011,918, doi: 10.1103/PhysRevE.80.011918
- Aihara I, Takeda R, Mizumoto T, Otsuka T, Takahashi T, Okuno HG, Aihara K (2011) Complex and transitive synchronization in a frustrated system of calling frogs. *Phys Rev E* 83(3):031,913, doi: 10.1103/PhysRevE.83.031913
- Asano F, Asoh H, Matsui T (1999) Sound source localization and signal separation for office robot “JiJo-2”. In: *Proc. of Multisensor Fusion and Integration for Intelligent Systems*, pp 243–248, doi: 10.1109/MFI.1999.815997
- Clark CW, Ellison WT (2000) Calibration and comparison of the acoustic location methods used during the spring migration of the bowhead whale, *balaena mysticetus*, off Pt. Barrow, Alaska, 1984–1993. *J of Acoust Soc Am* 107(6):3509–3517, doi: 10.1121/1.429421
- Farid H (2001) Blind inverse gamma correction. *IEEE Trans on Image Processing* 10(10):1428–1433, doi: 10.1109/83.951529
- Feng A, Narins PM, Xu CH, Lin WY, Yu ZL, Qiu Q, Z-M, Shen JX (2006) Ultrasonic communication in frogs. *Nature* 440:2333–2336, doi: 10.1038/nature04416
- Gerhardt H, Huber F (2002) *Acoustic communication in insects and anurans*. The University of Chicago Press, Chicago
- Grafe TU (1997) Costs and benefits of mate choice in the lek-breeding reed frog, *hyperolius marmoratus*. *Animal Behavior* 53:1103–1117, doi: 10.1006/anbe.1996.0427
- Hedwig B, Poulet J (2004) Complex auditory behavior emerges from simple reactive steering. *Nature* (430):781–785, doi: 10.1038/nature02787
- Hyvarinen A, Karhunen J, Oja E (2001) *Independent component analysis*. Wiley-Interscience, New York
- ITU-R (2007) Recommendation ITU-R BT.606-6: Studio encoding parameters of digital television for standard 4:3 and wide screen 16:9 aspect ratios. International Telecommunication Union Radiocommunication Sector

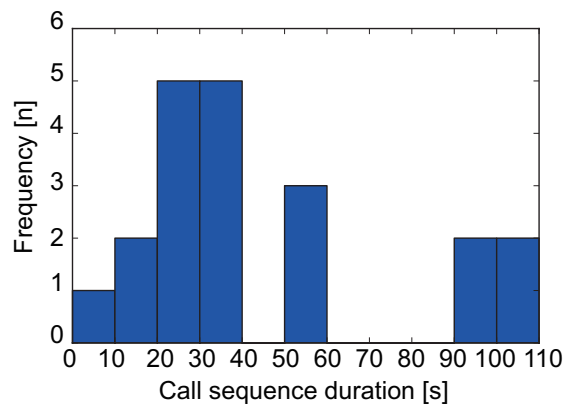
- Jones DL, Ratnam R (2009) Blind location and separation of callers in a natural chorus using a microphone array. *J of Acoust Soc Am* 126(2):895–910, doi: 10.1121/1.3158924
- Liu C (2005) Foundations of MEMS. Prentice Hall, New Jersey
- MacCurdy R, Fristrup K (2009) Automatic animal tracking using matched filters and time difference of arrival. *J of Communications* 4(7):487–495, doi: 10.4304/jcm.4.7.487-495
- Maeda N, Matsui M (1999) Frogs and toads of Japan, Bun-ichi Sogo Shuppan Co. Ltd., pp 36–39
- Matsui M (1996) Natural history of the amphibia, University of Tokyo Press, pp 150–152
- Nakadai K, Okuno HG, Nakajima H, Hasegawa Y, Tsujino H (2009) Design and implementation of robot audition system “HARK”. *Advanced Robotics* 24:739–761, doi: 10.1163/016918610X493561
- Nakajima H, Nakadai K, Hasegawa Y, Tsujino H (2010) Blind source separation with parameter-free adaptive step-size method for robot audition. *IEEE Trans on Audio, Speech, and Language Processing* 18(6):1476–1484, doi: 10.1109/TASL.2009.2035219
- Narins PM, Capranica RR (1978) Communicative significance of the two-note call of the treefrog *eleutherodactylus coqui*. *J Comp Physiol A* 127:1–9, doi: 10.1007/BF00611921
- Otsu N (1979) A threshold selection method from gray-level histograms. *IEEE Trans on Systems, Man, and Cybernetics* SMC-9(1):62–66, doi: 10.1109/TSMC.1979.4310076
- Riquimaroux H, Gaioni SJ, Suga N (1991) Cortical computational maps control the auditory perception. *Science* (251):565–568, doi: 10.1126/science.1990432
- Sawada H, Mukai R, Araki S (2003) Polar coordinate based nonlinear function for frequency-domain blind source separation. *IEICE Trans on Fundamentals of Electronics, Communications and Computer Sciences* 86(3):590–596
- Schwartz JJ (2001) Call monitoring and interactive playback systems in the study of acoustic interactions among male anurans, Smithsonian Institution Press, Washington and Chigaco, pp 183–204
- Simmons AM (2004) Call recognition in the bullfrog, *rana catesbiana*: Generalization along the duration continuum. *J of Acoust Soc Am* 115(3):1345–1355, doi: 10.1121/1.1643366
- Simmons AM, Simmons JA, Bates ME (2008) Analyzing acoustic interactions in natural bullfrog (*rana catesbeiana*) choruses. *J Comp Physiol* 122(3):274–282, doi: 10.1037/0735-7036.122.3.274
- Spiesberger JL (1999) Locating animals from their sounds and tomography of the atmosphere: Experimental demonstration. *J of Acoust Soc Am* 106(2):837–846, doi: 10.1121/1.427100
- Spiesberger JL, Fristrup KM (1990) Passive localization of calling animals and sensing of their acoustic environment using acoustic tomography. *American Naturalist* 135(1):107–153, doi: 10.1086/285035
- Suggs DN, Simmons AM (2005) Information theory analysis of patterns of modulation in the advertisement call of the male bullfrog, *rana catesbiana*. *J of Acoust Soc Am* 117(4):2330–2337, doi: 10.1121/1.1863693
- Wells K (2007) The ecology and behaviour of amphibians. The University of Chicago Press, Chicago



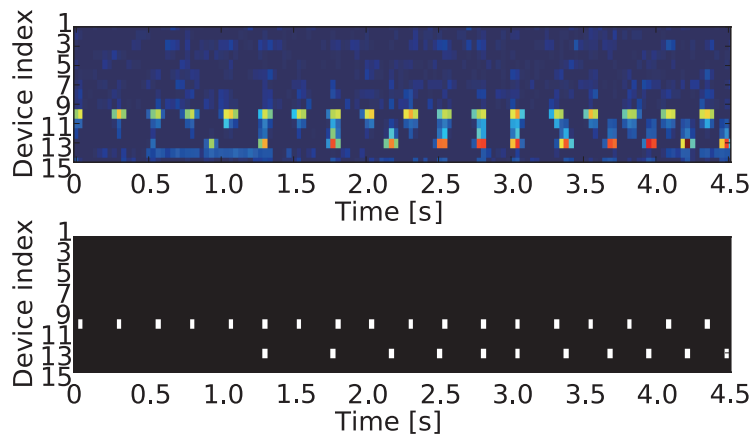
**Fig. 1 Experimental conditions:** In the indoor experiment, we used 10 Fireflies (rectangles) in each row at a 21.5cm interval, two frogs (gray circles), and a video camera placed overhead. In the field experiment, we deployed 20 Fireflies in a row along the edge of a rice paddy 6m long, and place the video camera at the corner. The frogs' locations were unknown; Figure 1c shows an example of an image recorded.



**Fig. 2 Visualized spatio-temporal calling behavior of two frogs:** Figure 2a shows the complete visualization of calling behaviors for the indoor experiment. The x-, y-, and z-axes denote time, the device index, and the normalized intensity, i.e., the sound intensity of the call, respectively. The upper panel in Fig. 2b is a magnification of Fig. 2a from 15.0 to 17.0 s. The lower panel shows the localization; each white pixel represents a call. The anti-phase synchronized calling behavior is evident. Figure 2c shows the waveforms corresponding to the calls; they were separated using independent component analysis. Figures 2d and 2e show the corresponding results for the field experiment.



**Fig. 3 Histogram of the call sequence durations:** Histogram of call sequence durations observed in the field experiment. Each duration is the difference between the starting time of the first call in a sequence and ending time of the last call in the same sequence. The typical duration was between 20 and 40s, and the longest was 108s.



**Fig. 4 Visualization example for temporal overlapping:** The axis labels are the same as in Figs. 2b and 2e. The data are from the field experiment. From 1.0s to 3.0s, the system separately localized the overlapping calls of two frogs.